

Supplemental materials for the manuscript “Genome analysis of the fruiting body forming myxobacterium *Chondromyces crocatus* reveals high potential for natural product biosynthesis”

File 1 (of 2)

Nestor Zaburannyi^{1,4}, Boyke Bunk^{2,4}, Josef Maier^{3,5}, Jörg Overmann^{2,4}, Rolf Müller^{1,4#}

¹Department of Microbial Natural Products, Helmholtz Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Centre for Infection Research and Pharmaceutical Biotechnology, Saarland University, Saarbrücken, Germany.

²Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany.

³ATG-biosynthetics GmbH, Merzhausen, Germany

⁴German Center for Infection Research (DZIF), partner site Hannover-Braunschweig, Braunschweig, Germany.

⁵IStLS, Information Services to Life Science, Oberndorf a.N., Germany

Table of content:		Page
Fig. S1.	Non-cumulative GC-skew of the <i>Chondromyces crocatus</i> Cm c5	2
Fig. S2.	Leucine codons distribution in protein coding sequences of <i>C. crocatus</i> Cm	2
Fig. S3.	Growth of <i>C. crocatus</i> Cm c5 fr- and fr+ strains on hard and soft agar plates	3
Fig. S4.	Multiple sequence alignment and domain organization of NrdR proteins	4
Fig. S5.	Multiple sequence alignment and domain organization of ExbD proteins	5
Fig. S6.	Similarity of natural products linking secondary metabolisms of <i>C. crocatus</i> Cm c5 and sponges	5
Table S1.	Measures for bias in codon usage expressed as differences in medians of adaptation indices	6
Table S2.	List of eukaryotic-like serine/threonine protein kinases encoded in the genome of <i>C. crocatus</i> Cm c5.	7
Table S3.	Genes for Social (S-)motility	12
Table S4.	List of candidate horizontally transferred genes predicted by homology-phylogenetic approach	13
Table S5.	Secondary metabolite biosynthetic gene clusters	18
References used in Supplemental Materials		19

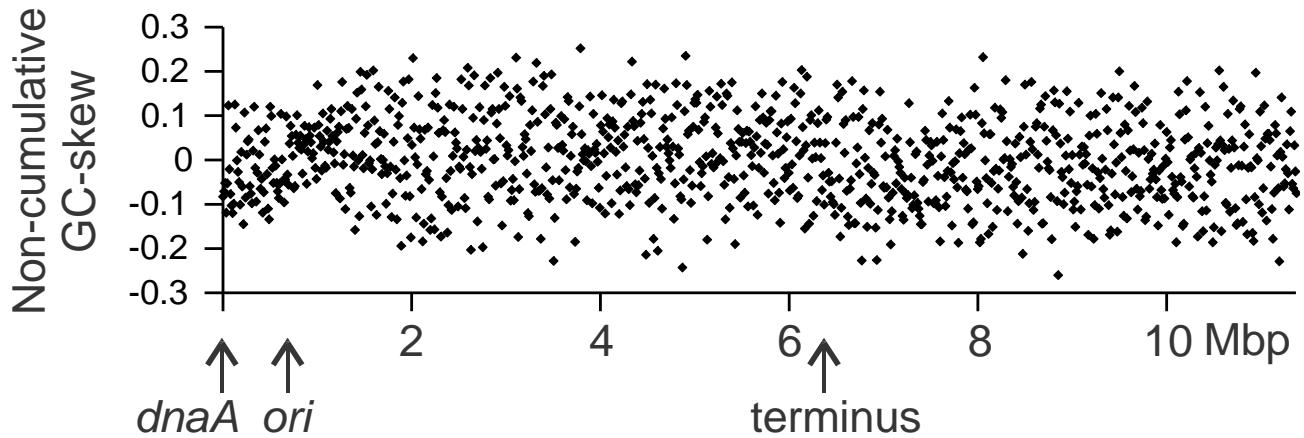


Fig. S1. Non-cumulative GC-skew of the *Chondromyces crocatus* Cm c5 chromosome. Each dot represents a window length of 1/1000 of the chromosome. Arrows point at the locations of the *dnaA* gene as well as *origin* and *terminus* of replication.

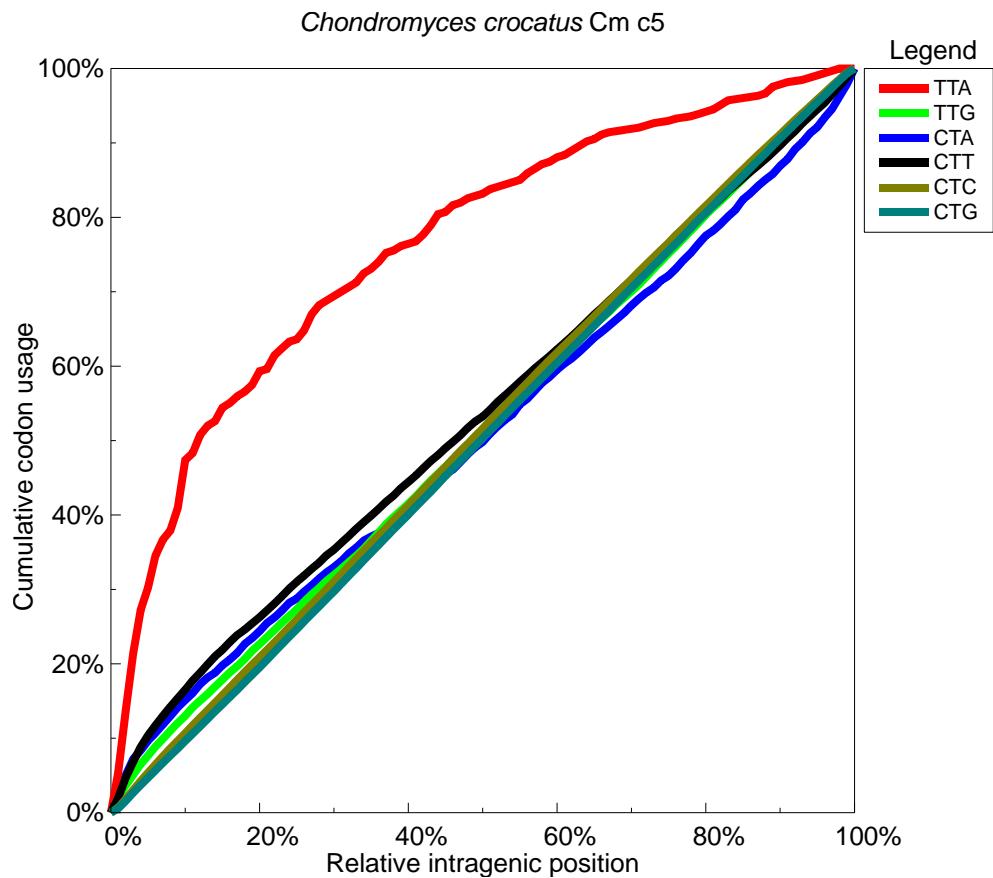


Fig. S2. Leucine codons distribution in protein coding sequences of *C. crocatus* Cm c5. Relative intragenic position withing coding sequences is plotted against the whole-genome codon usage (cumulative). Correlation coefficients when comparing to the theoretical uniform codon distribution are as follows: TTA: R=0.915; TTG: R>0,999; CTA: R>0,999; CTT: R>0,999; CTC: R>0,999; CTG: R>0,999.

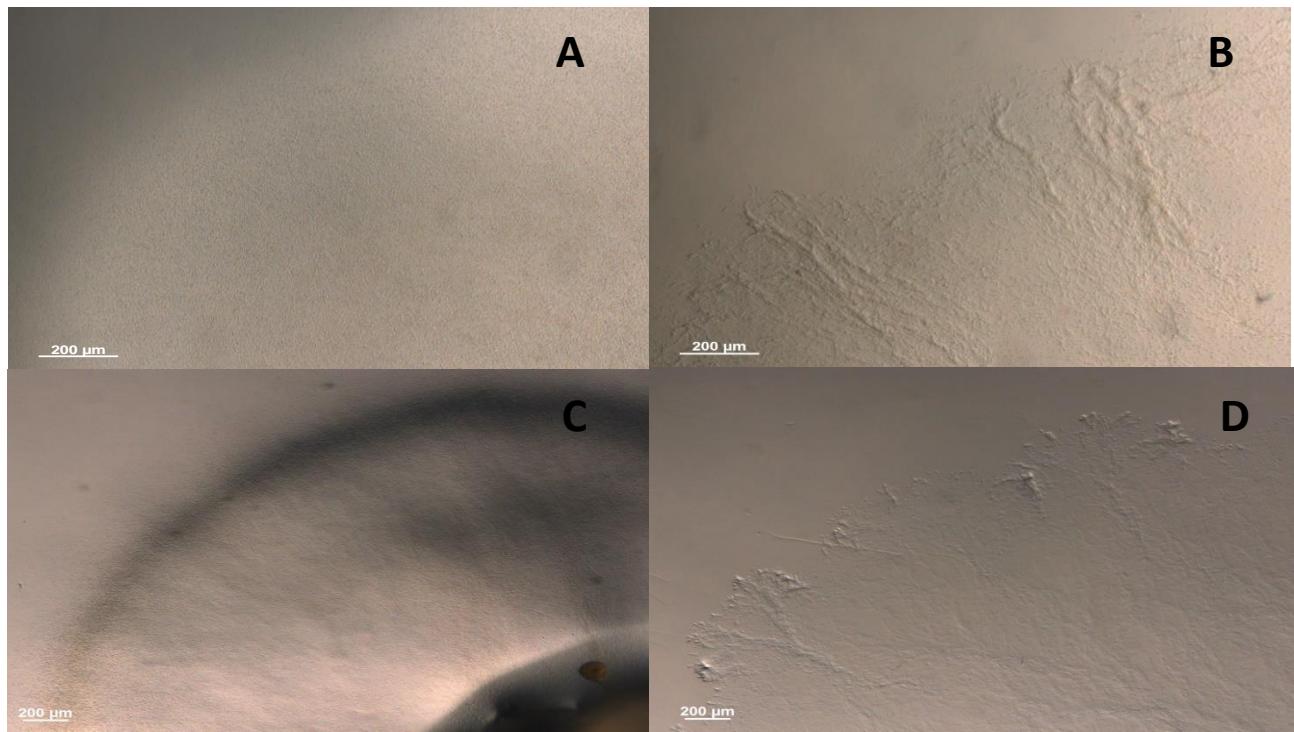


Fig. S3. Growth of *C. crocatus* Cm c5 fr- and fr+ strains on hard and soft agar plates. (A) Cm c5 fr- on 8g/L Bacto Agar; (B) Cm c5 fr- on 15g/L Bacto Agar; (C) Cm c5 fr+ on 8g/L Bacto Agar; (D) Cm c5 fr+ on 15g/L Bacto Agar;

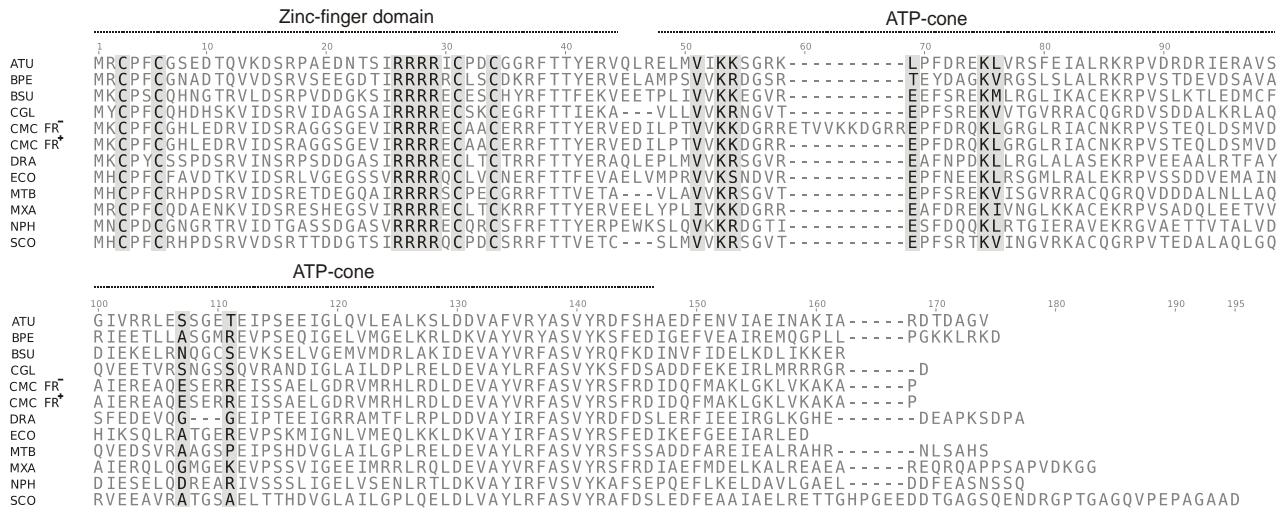


Fig. S4. Multiple sequence alignment and domain organization of NrdR proteins. Protein sequences were selected from a representative set of major groups of bacteria. Zinc-finger domain (positions 1-44) and ATP-cone (positions 48-146) shown as suggested in (1). Conserved residues that define the zinc finger motif and residues in ATP-cone presumably involved in nucleotide binding are highlighted. Abbreviations: ATU, *Agrobacterium tumefaciens*; BPE, *Bordetella pertussis*; BSU, *Bacillus subtilis*; CGL, *Corynebacterium glutamicum*; CMC, *Chondromyces crocatus*; DRA, *Deinococcus radiodurans*; ECO, *Escherichia coli*; MTU, *Mycobacterium tuberculosis*; MXA, *Myxococcus xanthus*; NPH, *Natronomonas pharaonis*; TMA, *Thermotoga maritima*; SCO, *Streptomyces coelicolor*; Protein sequences of NrdR used in multiple alignment were downloaded from NCBI GenBank database for the following organisms (accession codes): *Streptomyces coelicolor* A3(2) (NP_629928), *Mycobacterium tuberculosis* H37Rv (NP_217234), *Corynebacterium glutamicum* ATCC 13032 (NP_601134), *Bacillus subtilis* 168 (NP_390778), *Myxococcus xanthus* DK1622 (YP_632927), *Thermotoga maritima* MSB8 (AGL50639), *Deinococcus radiodurans* R1 (NP_293813), *Bordetella pertussis* Tohama I (NP_881530), *Agrobacterium tumefaciens* F2 (WP_003496675), *Natronomonas pharaonis* Gabara (WP_011324460), *Escherichia coli* str. K-12 substr. MG1655 (NP_414947).

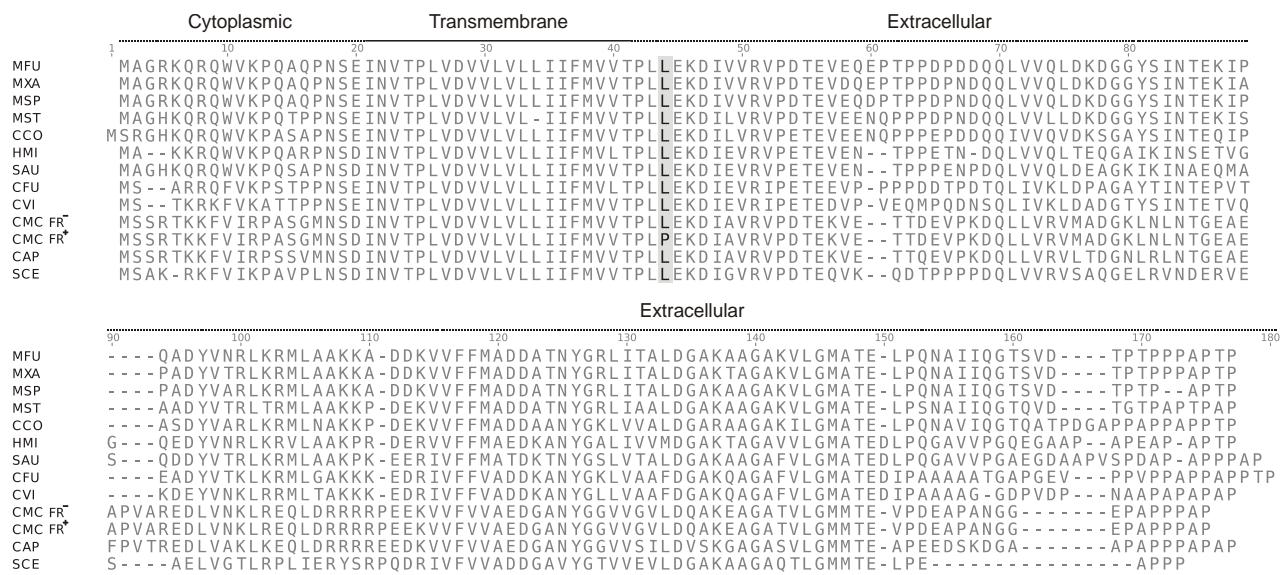


Fig. S5. Multiple sequence alignment and domain organization of ExbD proteins. Protein sequences were selected from a representative set of myxobacteria. Cytoplasmic domain (positions 1-20), transmembrane domain (positions 21-41) and extracellular domain (positions 42-164) shown as detected by Geneious software (2). Mutated leucine/proline residue is highlighted. Abbreviations: CAP, *Chondromyces apiculatus*; CCO, *Corallococcus coralloides*; CFU, *Cystobacter fuscus*; CMC, *Chondromyces crocatus*; CVI, *Cystobacter violaceus*; HMI, *Hyalangium minutum*; MFU, *Myxococcus fulvus*; MSP, *Myxococcus* sp. (contaminant ex DSM 436); MST, *Myxococcus stipitatus*; MXA, *Myxococcus xanthus*; SAU, *Stigmatella aurantiaca*; SCE, *Sorangium cellulosum*. Protein sequences of ExbD used in multiple alignment were downloaded from NCBI GenBank database for the following organisms (accession codes): *Chondromyces apiculatus* DSM 436 (EYF03710), *Cystobacter violaceus* Cb vi76 (KFA92750), *Hyalangium minutum* DSM 14724 (KFE67694), *Stigmatella aurantiaca* DW4/3-1 (WP_002610411), *Cystobacter fuscus* (WP_002629175), *Myxococcus* sp. (contaminant ex DSM 436) (WP_002635514), *Myxococcus xanthus* DK 1622 (WP_011551562), *Sorangium cellulosum* So ce56 (WP_012236884), *Myxococcus fulvus* HW-1 (WP_013935406), *Corallococcus coralloides* DSM 2259 (WP_014394197), *Myxococcus stipitatus* DSM 14675 (WP_015347076).

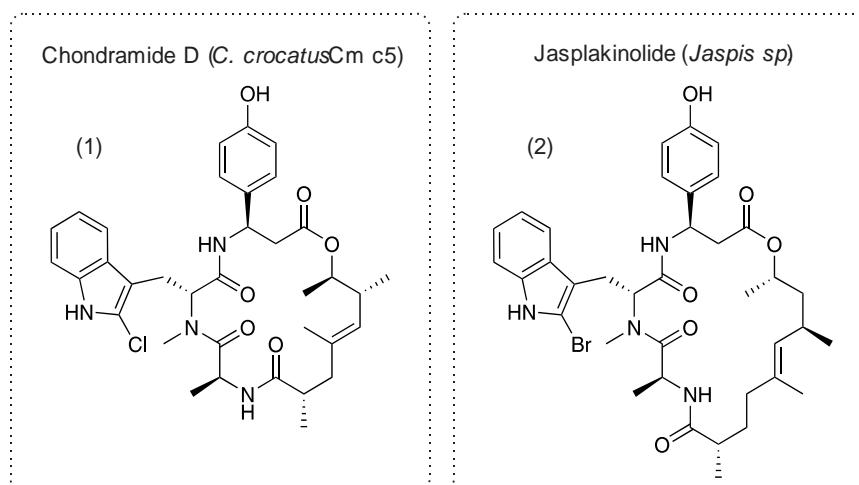


Fig. S6. Similarity of natural products linking secondary metabolisms of *C. crocatus* Cm c5 and sponges. (1) – chondramide D from *C. crocatus* Cm c5; (2) – jasplakinolide from *Jaspis* sp.

Table S1. Measures for bias in codon usage expressed as differences in medians of codon adaptation indices. For all CDS ($mCAI_{All}$), the presumably highly expressed Hxp2 CDS set ($mCAI_{Hxp2}$) and the megasynthase CDS set ($mCAI_{Msy}$), where the CAI was calculated with the synonymous codon usage table of all CDS (All) or of the Hxp2 CDS set (Hxp2), and the codon enrichment score ΔH per codon and its log likelihood ratio test statistic G, which indicates the strength of translational selection on codon bias. CUT, synonymous codon usage table. For calculation of p-values see Methods.

Genome	<i>C. crocatus</i> Cm c5	<i>S. cellulosum</i> So ce56	<i>M. xanthus</i> DK 1622	<i>E. coli</i> K12 MG1655			
CUT	All	Hxp2	All	Hxp2	All	Hxp2	All
n (CDS)	8339	351	9375	388	7331	353	4140
$mCAI_{All}$	0.739	0.683	0.734	0.671	0.716	0.637	0.711
$mCAI_{Hxp2}$	0.807	0.763	0.816	0.771	0.806	0.768	0.776
$\Delta mCAI_{Hxp2-All}$	0.068	0.080	0.082	0.100	0.090	0.131	0.065
P-value	<2.2E-16	<2.2E-16	<2.2E-16	<2.2E-16	<2.2E-16	<2.2E-16	<2.2E-16
ΔH_{Hxp2}	0.013		0.012		0.023		0.095
G_{Hxp2}	3976		3771		6383		16091
P-value _{Hxp2}	~0		~0		~0		~0
n (CDS Msy)	69		26		53		
$mCAI_{Msy}$	0.724	0.667	0.746	0.684	0.676	0.583	
$\Delta mCAI_{Msy-All}$	-0.015	-0.016	0.012	0.013	-0.040	-0.054	
P-value _{Msy-All}	0.289	0.334	0.261	0.257	0.004	0.003	
$\Delta mCAI_{Msy-Hxp2}$	-0.083	-0.096	-0.070	-0.087	-0.130	-0.185	
P-value _{Msy-Hxp2}	<2.2E-16	<2.2E-16	<2.2E-16	<2.2E-16	<2.2E-16	<2.2E-16	
ΔH_{Msy}	0.001		0.003		0.008		
G_{Msy}	442		460		2481		
P-value _{Msy}	1.12E-68		3.05E-72		~0		

Table S2. List of eukaryotic-like serine/threonine protein kinases encoded in the genome of *C. crocatus* Cm c5. Most common five domains types are shown as dots. The number of dots indicates the domain count within the respective coding sequences.

Locus	Length	Domains detected					
		PKc	ATPase	WD40	TPR	PEGA	Other domains
CMC5_000020	1308	•	•				
CMC5_000330	632	•					
CMC5_000170	1767	•	•				GAF, HisKA, HATPase_c
CMC5_001570	1027	• ¹					
CMC5_002330	858	•					
CMC5_002440	334	•					
CMC5_002710	697	•					
CMC5_003330	696	•					
CMC5_004280	1270	•	•				
CMC5_004760	661	•					
CMC5_005010	656	•					
CMC5_005390	447	•					
CMC5_006080	602	•					
CMC5_006340	538	• ¹					MSP1_C ¹
CMC5_006660	450	•					
CMC5_006740	575	•					
CMC5_007000	421	• ¹					
CMC5_007730	562	•					
CMC5_007760	571	•					
CMC5_007900	705	•					
CMC5_008060	1417	•			• ¹		
CMC5_008070	482	•					
CMC5_008160	573	•				• ¹	
CMC5_009550	554	•					Macro_Appr_pase_like ¹
CMC5_009890	756	•					
CMC5_009900	1036	•					
CMC5_010190	1094	• ¹					PDZ_serine
CMC5_010300	1310	•	•				
CMC5_010560	400	•					
CMC5_010820	611	•					
CMC5_010830	287	•					
CMC5_010860	545	•					
CMC5_011050	358	•					
CMC5_011330	618	• ¹					
CMC5_011770	321	•					
CMC5_011780	557	•				• ¹	
CMC5_011870	1309	•	•				
CMC5_012310	437	• ¹					
CMC5_012780	544	• ¹					
CMC5_012850	422	•					
CMC5_012950	666	• ¹					REC, CHD
CMC5_013100	1315	•	•				
CMC5_013120	1307	•	•				

CMC5_013520	391	•						
CMC5_013800	623	•						2TM
CMC5_013820	566	•						
CMC5_014290	520	•						
CMC5_014650	605	•						
CMC5_014660	521	•						
CMC5_014680	591	•						
CMC5_015070	598	•						
CMC5_015170	574	•						
CMC5_015350	601	•						zf-
CMC5_015590	491	•						
CMC5_015630	489	•						
CMC5_016210	403	•						
CMC5_016890	527	•						
CMC5_017370	1295	•	•		•			
CMC5_017930	486	•						
CMC5_018730	1361	•• ¹						
CMC5_018770	469	• ¹						
CMC5_018980	937	•						
CMC5_019620	439	• ¹						
CMC5_020080	591	•						
CMC5_020210	425	•						
CMC5_020230	710	•						
CMC5_020600	504	•						
CMC5_020870	461	•				• ¹		
CMC5_021030	751	•						
CMC5_021440	621	•						
CMC5_021800	617	•						
CMC5_022390	727	•						
CMC5_022900	1311	• ¹	•		•			
CMC5_022910	622	•						
CMC5_023020	511	•						
CMC5_024450	816	•						FGE-sulfatase ¹
CMC5_024750	1267	•	•					
CMC5_025000	650	•						
CMC5_025320	603	•						
CMC5_025380	657	•						REC, CHD
CMC5_025820	465	•						REC
CMC5_025860	504	•						PRX_BCP
CMC5_025920	282	• ¹						
CMC5_026030	356	•						
CMC5_026040	638	•						
CMC5_026490	1412	• ¹	•					
CMC5_026930	675	•						
CMC5_027130	558	•						
CMC5_028120	722	•						FGE-sulfatase ¹
CMC5_028810	581	•						
CMC5_028840	565	•						FHA

CMC5_029060	581	•						
CMC5_029540	462	•						
CMC5_029630	524	•						
CMC5_029670	483	•						
CMC5_029770	355	•						
CMC5_030710	448	• ¹						
CMC5_030950	504	•						
CMC5_031770	708	•						
CMC5_031930	1667	•						
CMC5_032210	285	•						
CMC5_032430	537	•						
CMC5_033040	576	•						
CMC5_033430	595	•						
CMC5_033560	605	•						zf-, FGE-sulfatase ¹
CMC5_033610	329	•						
CMC5_034010	522	•						
CMC5_034240	655	• ¹						
CMC5_034450	533	•						
CMC5_035370	728	•						
CMC5_036050	1312	•	•			•		
CMC5_036630	895	•				•• ¹		
CMC5_036940	684	•						
CMC5_037790	652	•						
CMC5_037920	542	•						
CMC5_037970	689	•						Pup_ligase
CMC5_039360	632	• ¹						
CMC5_039400	927	•	•					zf- x 2
CMC5_039900	556	•						
CMC5_040350	1320	•	•					
CMC5_040380	879	• ¹	•					
CMC5_040420	324	•						
CMC5_040850	532	•						
CMC5_041020	553	•						
CMC5_041330	446	•						
CMC5_041440	1205	• ¹	•					
CMC5_042040	1301	•	•			•		
CMC5_042190	1170	•	•					
CMC5_042220	293	•						
CMC5_042650	911	•						LanC_like
CMC5_043390	1198	• ¹						
CMC5_043650	864	•						
CMC5_044240	439	• ¹						
CMC5_044310	444	• ¹						
CMC5_044680	1324	• ¹	•					
CMC5_044730	561	•						FHA
CMC5_044760	541	•						
CMC5_044980	559	•						
CMC5_045120	574	•						

CMC5_045400	568	•						
CMC5_045960	460	• ¹						
CMC5_046940	466	•						
CMC5_047090	1127	•	••					
CMC5_048410	558	•						
CMC5_048770	529	•						
CMC5_050020	550	• ¹						
CMC5_050220	420	•						
CMC5_050250	463	• ¹						
CMC5_050700	546	•						
CMC5_051230	577	•						
CMC5_051260	468	• ¹						
CMC5_051990	474	•						
CMC5_052330	1323	• ¹	•			• ¹	FGE-sulfatase ¹	
CMC5_052490	623	•						
CMC5_052800	650	•						
CMC5_052810	628	•						
CMC5_053900	566	•						
CMC5_056140	583	•						
CMC5_056170	582	•						
CMC5_056240	954	••						
CMC5_056870	1190	•	•					
CMC5_056990	447	• ¹						
CMC5_057550	1539	• ¹	•					
CMC5_057570	550	•						
CMC5_057890	1576	•		••••				
CMC5_058280	1315	•	•					
CMC5_058310	594	•						
CMC5_059000	418	•						
CMC5_059250	576	• ¹						
CMC5_059310	358	•						
CMC5_059770	514	•						
CMC5_060180	916	•			•• ¹			
CMC5_060550	527	•						
CMC5_060590	334	• ¹						
CMC5_060920	716	•						
CMC5_061600	886	•						
CMC5_062140	571	•						
CMC5_062660	1098	•						
CMC5_062860	577	•					Laminin_G_3	
CMC5_062870	805	•						
CMC5_063060	321	•						
CMC5_063700	570	•						
CMC5_063930	1158	•			•			
CMC5_064500	873	•					FGE-sulfatase ¹	
CMC5_064930	350	• ¹						
CMC5_065700	551	•						
CMC5_066020	587	•						

CMC5_066140	299	•						
CMC5_066150	692	•						
CMC5_066250	812	•						
CMC5_066260	738	•						
CMC5_066740	516	•						
CMC5_067830	574	•						
CMC5_067950	596	•						Pentapeptide repeat
CMC5_068530	1190	• ¹		•••				
CMC5_068710	501	•						
CMC5_068810	521	•						
CMC5_069030	495	•						
CMC5_069220	899	•	•					
CMC5_070130	843	•			•			
CMC5_070520	586	•						
CMC5_071510	929	• ¹						
CMC5_071810	953	•						
CMC5_071820	699	•						
CMC5_072100	1176	•						
CMC5_072210	468	•						
CMC5_072880	630	•						
CMC5_073320	672	•						
CMC5_073410	623	•						
CMC5_073480	349	•						
CMC5_073510	587	•						
CMC5_074620	547	•						
CMC5_074630	593	• ¹						
CMC5_075140	423	•						
CMC5_075170	1822	•	•					GAF, HisKA, HATPase_c
CMC5_075210	447	•						
CMC5_075660	707	•						HYR
CMC5_075840	353	•						
CMC5_075870	919	•						GH_43_62_32_68 ¹
CMC5_076270	917	•				• ¹		FGE-sulfatase ¹
CMC5_076660	855	•						
CMC5_076940	1051	• ¹						
CMC5_077480	499	•						
CMC5_077660	530	• ¹						
CMC5_077740	900	•						
CMC5_078850	459	•						
CMC5_078870	1355	•			•			
CMC5_079020	1432	• ¹	•		•			
CMC5_079120	431	•						
CMC5_079380	576	•						
CMC5_079400	795	•						
CMC5_079420	565	•						
CMC5_079530	480	•						REC
CMC5_079550	937	• ¹			•			
CMC5_079720	552	• ¹						

CMC5_079970	269	•						
CMC5_080250	855	•				• ¹	FGE-sulfatase ¹	
CMC5_080610	628	•				• ¹		
CMC5_081090	578	•						
CMC5_081500	676	•						
CMC5_081800	1353	•	•		•			
CMC5_082460	489	•					PRK00431 ¹	
CMC5_082800	584	•						
CMC5_082930	1566	•	•	•••				
CMC5_083420	491	•						
CMC5_083520	558	•						
CMC5_083980	317	•						
CMC5_083990	486	• ¹						
CMC5_084120	518	•					REC	
CMC5_084430	621	•						
Total: 250 genes								

¹superfamily prediction only

Note: Domain designations as used in Conserved Domains Database (3).

Table S3. Genes for Social (S-)motility

Gene	Locus	Length (aa)	Function
pilB	CMC5_059020	566	type IV pilus biogenesis protein PilB (Pilus extension)
pilT	CMC5_059030	385	twitching motility protein PilT (Pilus retraction)
pilC	CMC5_024950	406	type IV pilin assembly protein PilC (Unknown)
pilS	Not Determined	-	pilin regulatory protein PilS (Two-component sensor for PilR)
pilR	CMC5_052090	506	pilin regulatory protein PilR (regulates transcription of pilA)
pilA1	CMC5_052100	196	fimbrial protein PilA1 (Pilin, monomer unit of the pilus filament)
pilA2	CMC5_052110	199	fimbrial protein PilA2 (Pilin, monomer unit of the pilus filament)
pilA3	CMC5_052120	204	fimbrial protein PilA3 (Pilin, monomer unit of the pilus filament)
pilD	CMC5_003390	322	Prepilin peptidase PilD (PilA leader peptidase)
pilM	CMC5_025260	353	Type IV pilus biogenesis protein PilM (ATPase)
pilN	CMC5_025270	213	Type IV pilus biogenesis protein PilN (Unknown)
pilO	CMC5_025280	219	Type IV pilus assembly protein PilO (Unknown)
pilP	CMC5_025290	233	Type IV pilus assembly protein PilP (Unknown)
pilQ	CMC5_025300	777	Type IV pilus biogenesis and competence protein PilQ (Secretin)
tgl	Not Determined	-	social gliding motility protein Tgl (Secretin assembly factor)

Table S4. List of candidate horizontally transferred genes predicted by homology-phylogenetic approach

Locus	Annotated product
CMC5_00055	photosystem I assembly BtpA
CMC5_00079	sulfate adenylyltransferase subunit 2
CMC5_00080	adenylylsulfate kinase
CMC5_00166	MFS transporter
CMC5_00174	MFS transporter
CMC5_00204	glycogen debranching protein
CMC5_00459	methionine sulfoxide reductase A
CMC5_00624	glutamate dehydrogenase
CMC5_00626	aminotransferase
CMC5_00710	leucine dehydrogenase
CMC5_00782	phenazine biosynthesis protein PhzF
CMC5_00826	radical SAM protein
CMC5_01047	hypothetical protein
CMC5_01072	multidrug transporter AcrB
CMC5_01128	molybdopterin biosynthesis protein MoeA
CMC5_01154	multidrug transporter
CMC5_01243	dioxygenase
CMC5_01251	glucose-1-phosphate thymidylyltransferase
CMC5_01252	dTDP-glucose 4,6-dehydratase
CMC5_01262	glycosyltransferase
CMC5_01265	sulfate ABC transporter
CMC5_01270	S-formylglutathione hydrolase
CMC5_01321	RNA 2'-phosphotransferase
CMC5_01341	pyridine nucleotide transhydrogenase
CMC5_01359	omega amino acid--pyruvate aminotransferase
CMC5_01581	glutamine amidotransferase
CMC5_01672	hypothetical protein
CMC5_01673	membrane protein
CMC5_01696	exopolyphosphatase
CMC5_01741	methionine ABC transporter permease
CMC5_01742	methionine ABC transporter substrate-binding protein
CMC5_01748	aldo/keto reductase
CMC5_01797	phosphate transporter PitA
CMC5_01809	nitroreductase
CMC5_01812	cation:proton antiporter
CMC5_01813	cation:proton antiporter
CMC5_01814	cation:proton antiporter
CMC5_01815	cation:proton antiporter
CMC5_01816	cation:proton antiporter

CMC5_01817	cation:proton antiporter
CMC5_01828	50S ribosomal protein L21
CMC5_01923	UDP-glucose 4-epimerase
CMC5_01967	glutamine amidotransferase
CMC5_01982	transposase
CMC5_02092	isocitrate dehydrogenase
CMC5_02118	C4-dicarboxylate ABC transporter
CMC5_02259	LysR family transcriptional regulator
CMC5_02276	hypothetical protein
CMC5_02316	alcohol dehydrogenase
CMC5_02364	short-chain dehydrogenase
CMC5_02369	short-chain fatty acid transporter
CMC5_02488	5-methyltetrahydropteroylglutamate-- homocysteine methyltransferase
CMC5_02505	ethanolamine ammonia-lyase small subunit
CMC5_02506	ethanolamine ammonia lyase large subunit
CMC5_02540	dienelactone hydrolase
CMC5_02549	TetR family transcriptional regulator
CMC5_02559	carbamoyltransferase
CMC5_02572	4-hydroxybenzoate--CoA ligase
CMC5_02574	aldehyde dehydrogenase
CMC5_02575	transcriptional regulator
CMC5_02576	crotonase
CMC5_02577	benzoyl-CoA oxygenase
CMC5_02578	benzoyl-CoA oxygenase
CMC5_02596	glyoxalase
CMC5_02611	uracil-DNA glycosylase
CMC5_02692	12-oxophytodienoate reductase
CMC5_02704	DSBA oxidoreductase
CMC5_02788	short-chain dehydrogenase
CMC5_02810	Rrf2 family transcriptional regulator
CMC5_02811	dihydropteridine reductase
CMC5_02848	NmrA family protein
CMC5_02915	phosphoglyceromutase
CMC5_02929	7-cyano-7-deazaguanine reductase
CMC5_03125	DNA ligase
CMC5_03261	ATPase AAA
CMC5_03384	multidrug ABC transporter substrate-binding protein
CMC5_03385	ABC transporter
CMC5_03395	MerR family transcriptional regulator
CMC5_03417	serine/threonine dehydratase
CMC5_03484	tryptophanyl-tRNA synthetase
CMC5_03516	membrane protein
CMC5_03549	thiol peroxidase

CMC5_03573	iron ABC transporter
CMC5_03617	aldo/keto reductase
CMC5_03624	ferredoxin reductase
CMC5_03738	3-oxoacyl-ACP reductase
CMC5_03860	acetyltransferase
CMC5_04165	glycerol-3-phosphatase
CMC5_04166	dehydrogenase
CMC5_04210	uridine/cytidine kinase
CMC5_04254	NADPH:quinone oxidoreductase
CMC5_04260	GTP-binding protein
CMC5_04282	hydroxylase
CMC5_04287	aspartate-semialdehyde dehydrogenase
CMC5_04306	argininosuccinate synthase
CMC5_04307	acetylornithine deacetylase
CMC5_04309	n-acetyl-gamma-glutamyl-phosphate reductase
CMC5_04310	argininosuccinate lyase
CMC5_04312	hypothetical protein
CMC5_04380	ABC transporter ATP-binding protein
CMC5_04387	6-phosphogluconate dehydrogenase
CMC5_04454	hypothetical protein
CMC5_04483	hypothetical protein
CMC5_04497	hypothetical protein
CMC5_04554	strictosidine synthase
CMC5_04685	polyketide cyclase
CMC5_04700	phosphotransferase
CMC5_04702	hypothetical protein
CMC5_04714	FAD-containing monooxygenase EthA
CMC5_04734	iron transporter
CMC5_04769	2-nitropropane dioxygenase
CMC5_04851	aminotransferase DegT
CMC5_04949	AraC family transcriptional regulator
CMC5_04961	hypothetical protein
CMC5_04963	alcohol dehydrogenase
CMC5_05129	hypothetical protein
CMC5_05132	NAD(P)H quinone oxidoreductase
CMC5_05134	thioesterase
CMC5_05169	N-(5'-phosphoribosyl)anthranilate isomerase
CMC5_05170	feruloyl-CoA synthase
CMC5_05178	GCN5 family acetyltransferase
CMC5_05179	hypothetical protein
CMC5_05252	acetyltransferase
CMC5_05253	thiazole biosynthesis protein ThjJ
CMC5_05301	nitrilotriacetate monooxygenase

CMC5_05318	hypothetical protein
CMC5_05332	acyl-CoA dehydrogenase
CMC5_05337	acetyl-CoA acetyltransferase
CMC5_05338	NAD dependent epimerase/dehydratase
CMC5_05339	LysR family transcriptional regulator
CMC5_05341	short-chain dehydrogenase
CMC5_05362	oxidoreductase
CMC5_05374	phosphoribosyltransferase
CMC5_05375	DeoR family transcriptional regulator
CMC5_05642	phosphate ABC transporter permease
CMC5_05655	hypothetical protein
CMC5_05666	oxidoreductase
CMC5_05712	hypothetical protein
CMC5_05730	aldehyde-activating protein
CMC5_05737	alkylhydroperoxidase
CMC5_05745	hypothetical protein
CMC5_05747	phosphoglycerate mutase
CMC5_05749	diguanylate cyclase
CMC5_05795	protein-L-isoaspartate O-methyltransferase
CMC5_05807	ferredoxin
CMC5_05879	amidohydrolase
CMC5_05944	arylesterase
CMC5_05945	membrane protein
CMC5_05948	membrane protein
CMC5_05967	sulfate ABC transporter substrate-binding protein
CMC5_05981	deoxyhypusine synthase
CMC5_06234	aliphatic nitrilase
CMC5_06334	pyridine nucleotide-disulfide oxidoreductase
CMC5_06335	deaminase
CMC5_06336	TetR family transcriptional regulator
CMC5_06351	hypothetical protein
CMC5_06453	oxidoreductase
CMC5_06584	3-demethylubiquinone-9 3-methyltransferase
CMC5_06605	3-oxoacyl-ACP synthase
CMC5_06686	RNA polymerase subunit sigma-24
CMC5_06689	ATPase
CMC5_06696	deaminase reductase
CMC5_06711	FMN reductase
CMC5_06717	DNA alkylation repair protein
CMC5_06737	NADPH:quinone oxidoreductase
CMC5_06764	TetR family transcriptional regulator
CMC5_06945	aldehyde dehydrogenase
CMC5_06987	ABC transporter

CMC5_06997	hypothetical protein
CMC5_07009	methionine sulfoxide reductase B
CMC5_07071	3-dehydroquinate dehydratase
CMC5_07135	stationary phase survival protein SurE
CMC5_07222	aldehyde dehydrogenase
CMC5_07282	hypothetical protein
CMC5_07295	AraC family transcriptional regulator
CMC5_07318	hypothetical protein
CMC5_07321	ATPase
CMC5_07360	acetylornithine deacetylase
CMC5_07363	LOG family protein
CMC5_07490	carbon starvation protein A
CMC5_07503	phenylalanyl-tRNA synthetase subunit alpha
CMC5_07523	glutathione S-transferase
CMC5_07547	hypothetical protein
CMC5_07552	carbohydrate-binding protein
CMC5_07649	hypothetical protein
CMC5_07772	oxidoreductase
CMC5_07828	short-chain dehydrogenase
CMC5_07848	hypothetical protein
CMC5_07858	AMP-dependent synthetase
CMC5_07888	serine/threonine protein phosphatase
CMC5_07923	chemotaxis protein CheY
CMC5_07998	serine hydrolase
CMC5_08012	dimethylallyltransferase
CMC5_08089	hypothetical protein
CMC5_08122	malonate decarboxylase subunit alpha
CMC5_08123	triphosphoribosyl-dephospho-CoA synthase
CMC5_08125	malonate decarboxylase subunit beta
CMC5_08126	malonate decarboxylase subunit gamma
CMC5_08140	virginiamycin B lyase
CMC5_08245	hypothetical protein
CMC5_08347	carbon monoxide dehydrogenase
CMC5_08392	acyl-CoA dehydrogenase
CMC5_08456	restriction endonuclease subunit M

Table S5. Secondary metabolite biosynthetic gene clusters

Cluster	Type	From	To	Product, if known
<u>Cluster 1</u>	Nrps-t1pk	4,265	58,312	
<u>Cluster 2</u>	Other	345,821	390,668	
<u>Cluster 3</u>	T3pk	750,571	791,677	
<u>Cluster 4</u>	Bacteriocin-nrps-t1pk	1,040,335	1,086,743	
<u>Cluster 5</u>	Terpene	1,897,735	1,918,673	
<u>Cluster 6</u>	Terpene	2,091,670	2,112,647	
<u>Cluster 7</u>	Nrps	2,328,095	2,411,827	
<u>Cluster 8</u>	Siderophore	3,119,692	3,134,272	
<u>Cluster 9</u>	T1pk	3,142,122	3,188,643	
<u>Cluster 10</u>	Terpene	3,643,831	3,664,772	
<u>Cluster 11</u>	Bacteriocin	3,703,806	3,715,602	
<u>Cluster 12</u>	Nrps-t1pk	3,930,755	4,019,178	Crocacin
<u>Cluster 13</u>	Bacteriocin-lantipeptide	4,409,234	4,431,398	
<u>Cluster 14</u>	Nrps-t1pk	4,797,039	4,847,593	
<u>Cluster 15</u>	Bacteriocin-nrps	5,583,796	5,643,479	Crocapeptin
<u>Cluster 16</u>	Lantipeptide	5,727,501	5,750,236	
<u>Cluster 17</u>	Other	5,743,463	5,790,836	
<u>Cluster 18</u>	Nrps-t1pk	6,371,337	6,445,775	Chondramide
<u>Cluster 19</u>	T1pk	6,542,182	6,613,442	
<u>Cluster 20</u>	T1pk	6,618,838	6,664,099	
<u>Cluster 21</u>	Lantipeptide-bacteriocin	6,930,080	6,960,482	
<u>Cluster 22</u>	T1pk-nrps	7,160,030	7,292,518	Chondrochloren
<u>Cluster 23</u>	Nrps-t1pk	7,274,640	7,348,387	
<u>Cluster 24</u>	T1pk-nrps	7,336,435	7,434,839	Thuggacin
<u>Cluster 25</u>	T1pk-nrps	7,478,242	7,589,080	Ajudazol
<u>Cluster 26</u>	Nrps-t1pk	7,618,352	7,666,156	
<u>Cluster 27</u>	Other	7,824,209	7,868,375	
<u>Cluster 28</u>	Phenazine	8,360,267	8,380,737	
<u>Cluster 29</u>	Nrps	9,030,294	9,085,101	
<u>Cluster 30</u>	Terpene	9,247,896	9,270,178	
<u>Cluster 31</u>	Terpene	9,508,549	9,529,457	
<u>Cluster 32</u>	Other	9,710,054	9,757,031	
<u>Cluster 33</u>	T1pk-nrps	9,743,580	9,793,738	
<u>Cluster 34</u>	T1pk	10,104,516	10,148,892	
<u>Cluster 35</u>	Bacteriocin	10,148,892	10,994,559	

References used in Supplemental Materials

1. **Grinberg I, Shtenberg T, Gorovitz B, Aharonowitz Y, Cohen G, Borovok I.** 2006. The Streptomyces NrdR transcriptional regulator is a Zn ribbon/ATP cone protein that binds to the promoter regions of class Ia and class II ribonucleotide reductase operons. *J Bacteriol* **188**:7635–7644.
2. **Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A.** 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**:1647–1649.
3. **Amorós-Moya D, Bedhomme S, Hermann M, Bravo IG.** 2010. Evolution in regulatory regions rapidly compensates the cost of nonoptimal codon usage. *Mol Biol Evol* **27**:2141–2151.